The University of Connecticut **Fall 2009**
Dept. of ESE **KRP**

Problem Set # 6

(Due Date: October 20, 2009)

1. Consider the problem of minimizing a quadratic function $f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} - \underline{x}^T \underline{b}$ where Q is symmetric and sparse (that is, there are relatively few nonzero entries in Q). The matrix Q has the form $Q = I + V$ where I is the identity matrix and V is a matrix with eigen values bounded by $e < 1$ in magnitude.
   a. With the given information, what is the best bound you can give for the rate of convergence of steepest descent applied to this problem?
   b. In general, it is difficult to invert Q, but the inverse can be approximated by $I - V$, which is easy to calculate (The approximation is very good for small e. Recall $(1-\varepsilon)^{-1} \approx 1+\varepsilon$ for small value of ε). We are thus led to consider the iterative process
   $$\underline{x}_{k+1} = \underline{x}_k - \alpha_k [I - V]\underline{g}_k$$
   where $\underline{g}_k = Q\underline{x}_k - \underline{b}$ and α_k is chosen to minimize f via line search. With the information given, what is the best bound on the rate of convergence of the method?
   c. Show that for $e < \frac{\sqrt{5}-1}{2} = 0.618$, the method in part (b) is always superior to steepest descent.
2. The following algorithm has been proposed for minimizing unconstrained functions $f(\underline{x})$ without using gradients: Starting with some arbitrary point $\underline{x}_0$, obtain a direction of search $\underline{d}_k$ such that for each component of $\underline{d}_k$
   $$f(\underline{x}_k + (\underline{d}_k)_i \underline{e}_i) = \min_{-\infty < d_i < \infty} f(\underline{x}_k + d_i \underline{e}_i),$$
   where $\underline{e}_i$ denotes the i^{th} column of the identity matrix. The next point $\underline{x}_{k+1}$ is determined the usual way via a line search along $\underline{d}_k$.
   a. Obtain an explicit representation of the algorithm for the quadratic case where $f(\underline{x}) = \frac{1}{2}(\underline{x} - \underline{x}^*)^T Q(\underline{x} - \underline{x}^*) + f(\underline{x}^*)$.
   b. Derive the convergence rate of this algorithm for the quadratic objective function. Express your answer in terms of the condition number of some matrix
   c. What condition on $f(\underline{x})$ or its derivatives will guarantee descent of this algorithm for general $f(\underline{x})$.
3. Show that if H_0 = I, the Davidon-Fletcher-Powell method is the conjugate gradient method. What similar statement can be made when H_0 is an arbitrary symmetric positive definite matrix?

4. Show that $\gamma_k = \dfrac{\underline{p}_k^T H_k^{-1} \underline{p}_k}{\underline{p}_k^T \underline{q}_k}$ also serves as a suitable scale factor for a self-scaling quasi-Newton method. Then, show that the convex combination $\gamma_k = (1-\alpha)\dfrac{\underline{p}_k^T \underline{q}_k}{\underline{q}_k^T H_k \underline{q}_k} + \alpha \dfrac{\underline{p}_k^T H_k^{-1} \underline{p}_k}{\underline{p}_k^T \underline{q}_k}$ for $0 \le \alpha \le 1$ is also a suitable scale factor.

5. (a) Show that the problem $\min\limits_{B \in R^{nxn}} \| B - A \|_F$ subject to B symmetric is solved by $B = \dfrac{1}{2}(A + A^T)$.

   (b) Prove that if Q is positive definite

$$\frac{\underline{p}^T \underline{p}}{\underline{p}^T Q \underline{p}} \le \frac{\underline{p}^T Q^{-1} \underline{p}}{\underline{p}^T \underline{p}}$$

   for any vector $\underline{p}$.

Computational (Due November 3, 2009):

Using the line search algorithm that you have developed, write a program for minimizing an arbitrary function $f(\underline{x}), \underline{x} \in R^n$ via the (i) diagonally-scaled steepest descent method, (ii) modified Newton methods (based on both step length and double dog leg curve methods), (iii) partial and pre-conditioned conjugate gradient methods, and (iv) quasi-Newton methods. The program should contain the following:

1. Input of n and starting point $\underline{x}_0$
2. Function (or subroutine) modules for evaluating $f(\underline{x}), \nabla \underline{f}(\underline{x})$, and $\nabla^2 f(\underline{x})$
3. Input of convergence tolerances for whatever convergence tests you decide to use
4. Keep count of the number of function evaluations
5. Input the line search tolerance parameter
6. Anything else you deem appropriate

Evaluate the program via application to:

1. Several quadratic functions having different condition numbers
2. Nonlinear functions of your choice

In the evaluation, examine convergence speed versus accuracy of line search. What happens if you do not reset or if reset with $m < n$. Your grade on this assignment depends on how well you demonstrate your knowledge of the unconstrained optimization methods, using as a vehicle the computer program.